On the Certification of the Restricted Isometry Property

Pascal Koiran* Anastasios Zouzias†

January 18, 2013

Abstract

Compressed sensing is a technique for finding sparse solutions to underdetermined linear systems. This technique relies on properties of the sensing matrix such as the *restricted isometry property*. Sensing matrices that satisfy this property with optimal parameters are mainly obtained via probabilistic arguments. Given any matrix, deciding whether it satisfies the restricted isometry property is a non-trivial computational problem. In this paper, we give reductions from dense subgraph problems to the certification of the restricted isometry property. This gives evidence that certifying this property is unlikely to be feasible in polynomial-time. Moreover, on the positive side we propose an improvement on the brute-force enumeration algorithm for checking the restricted isometry property.

Another contribution of independent interest is a spectral algorithm for certifying that a random graph does not contain any dense k-subgraph. This "skewed spectral algorithm" performs better than the basic spectral algorithm in a certain range of parameters.

1 Introduction

Let Φ be a $n \times N$ matrix with $N \geq n$. A vector $x \in \mathbb{C}^N$ is said to be k-sparse if it has at most k nonzero coordinates. Given $\delta \in]0,1[$, ϕ is said to satisfy the Restricted Isometry Property (RIP) of order k with parameter δ if it approximately preserves the Euclidean norm in the following sense: for every k-sparse vector x, we have

$$(1-\delta)||x||^2 \leq ||\Phi x||^2 \leq (1+\delta)||x||^2.$$

Clearly, for this to be possible we must have $k \leq n$. Given δ, n and N, the goal is to construct RIP matrices with k as large as possible. This problem is motivated by its applications to compressed sensing: it is known from Candès, Romberg and Tao [8, 9, 10] that the restricted isometry property enables the efficient recovery of sparse signals using linear programming techniques. For that purpose one can take any fixed $\delta < \sqrt{2} - 1$ [8].

Various probabilistic models are known to generate random matrices that satisfy the RIP with a value of k which is (almost) linear n. See for instance Theorem 2 in Section 3 for the case of matrices with entries that are independent symmetric (± 1) Bernouilli matrices. The recent survey [21] provides additional results of this type and extensive references to the probabilistic

*LIP, UMR 5668, ENS de Lyon – CNRS – UCBL – INRIA, École Normale Supérieure de Lyon, Université de Lyon and Department of Computer Science, University of Toronto, Canada Pascal.Koiran@ens-lyon.fr

†Department of Computer Science, University of Toronto, Canada zouzias@cs.toronto.edu

literature. Some significant effort has been devoted to the construction of explicit (rather than probabilistic) RIP matrices, but this appears to be a difficult problem. As pointed out by Bourgain et al. in a recent paper [6, 7], most of the known explicit constructions [17, 2, 13] are based on the construction of systems of unit vectors with a small coherence parameter (see section 2 for a definition of this parameter and its connection to the RIP). Unfortunately, this method cannot produce RIP matrices of order $k > \sqrt{n}$ [6, 7]. Bourgain et al. still manage to break through the $\sqrt{n}$ "barrier" using techniques from additive combinatorics: they construct RIP matrices of order $k = n^{1/2+\epsilon_0}$ where $\epsilon_0 > 0$ is an unspecified "explicit constant". Note that this is still far from the order achieved by probabilistic constructions.

Here we study the restricted isometry property from the point of view of computational complexity: what is the complexity of deciding whether a matrix satisfies the RIP, and of computing or approximating its order k or its RIP parameter δ? An efficient (deterministic) algorithm would have applications to the construction of RIP matrices. One would draw a random matrix Φ from one of the well-established probabilistic models mentioned above, and run this hypothetical algorithm on Φ to compute or approximate k and δ. The result would be a matrix with *certified* restricted isometry properties (see Section 3 for an actual result along those lines). This may be the next best thing short of an explicit construction (and as mentioned above, the known explicit constructions are far from optimal).

The definition of the restricted isometry property suggests an exhaustive search over $\binom{N}{k}$ subspaces, but we are not aware of any existing hardness result suggesting that exhaustive search is somehow unavoidable. There has been more work from the algorithm design side. In particular, it was shown that semi-definite programming can be used to verify the restricted isometry property [11] and other related properties from compressed sensing [12, 16]. Unfortunately, as pointed out in [12] these methods are unable to certify the restricted isometry property for k larger than $O(\sqrt{n})$, even for matrices that satisfy the RIP up to order $\Omega(n)$. As we have seen, $k = O(\sqrt{n})$ is also the range where coherence-based methods reach their limits.

In this paper we provide both positive and negative results on the computational complexity of the RIP, including the range $k > \sqrt{n}$.

Positive Results

In Section 2, we study the relation between the RIP parameters of different orders for a given matrix Φ. Very roughly, we show in Theorem 1 that the RIP parameter is at most proportional to the order. We therefore have a trade-off between order and RIP parameter: in order to construct a matrix of given order and RIP parameter, it suffices to construct a matrix of lower order and smaller RIP parameter. We illustrate this point in Section 3. Our starting point is the above-mentioned (very naive) exhaustive search algorithm, which enumerates all $\binom{N}{k}$ subspaces generated by k column vectors. We obtain a "lazy algorithm" which enumerates instead all subspaces generated by l basis vectors for some $l < k$. We show that the lazy algorithm can go slightly beyond the $\sqrt{n}$ barrier if a quasi-polynomial running time is allowed.

Negative Results

We provide the first hardness results on the certification of the restricted isometry property. We could not base these results on a well-established assumption such as $\mathsf{P} \neq \mathsf{NP}$. Instead, we rely on certain assumptions on the complexity of dense subgraph problems. There is already a fairly large

literature on the complexity of detecting dense subgraphs in an input graph, with both algorithmic results and hardness results (see [5] for extensive references). In particular, it was shown by Khot [18] that the dense k-subgraph problem does not have a polynomial time approximation scheme under the assumption that NP does not have randomized algorithms that run in sub-exponential time. Earlier, Feige [14] had obtained this result under the assumption that random 3-SAT formulas are hard to refute. Our results are of a similar flavor as those of Feige [14] and Alekhnovich [1] since they are based on average-case assumptions. However, dense subgraphs do not appear at the end of the reduction as in [14] but at the beginning. Our starting point is the assumption that it is hard to certify that a random $G(n, 1/2)$ graph does not contain any k-subgraph with density at least $1/2 + \epsilon$ (note that $1/2$ is the expected density). A certification algorithm must certify most of the graphs on n vertices, and any certified graph cannot contain any k-subgraph with density at least $1/2 + \epsilon$. More precisely, we present two distinct hardness hypotheses corresponding to different settings for the parameters k and ϵ. In the first hypothesis we set $k = n^\alpha$ and $\epsilon = 1/n^\beta$, where $\alpha, \beta \in]0, 1[$ and $\alpha - \beta < 1/2$. In the second hypothesis we set $k = \alpha n$ and $\epsilon = \gamma/\sqrt{n}$ where $\alpha \in]0, 1[$ is a small enough constant, $\gamma > 0$ and $\alpha\gamma$ is small enough. Here, "small enough" means that α and $\alpha\gamma$ are both upper bounded by κ, where $\kappa < 1$ is a universal constant. The second hypothesis postulates that with the above parameter settings, there is a value of κ for which certifying most graphs on n vertices is computationally difficult.

We believe that these two hypotheses are consistent with the current state of knowledge on dense subgraph problems. In particular, without the constraints on α, β and γ these hypotheses would be refuted by a simple spectral algorithm. Each of the two hypotheses leads to different hardness results. From the first hypothesis it follows that RIP parameters cannot be approximated to within any constant factor in polynomial time. From the second hypothesis we derive the following stronger result: no polynomial time algorithm can distinguish between a matrix which satisfies the RIP of order k with parameter (say) $\kappa/1000$ and a matrix which does *not* satisfy the RIP with parameter $\kappa/2$ (here the constant 1000 can be replaced by any constant larger than 2).

Finally, in an effort to further test the validity of these two hypotheses we propose a *skewed spectral algorithm* for certifying the absence of a k-subgraph of density at least $1/2 + \epsilon$ in a random graph. This algorithm performs better than the basic spectral algorithm for certain settings of k and ϵ, but is still consistent with our two hypotheses.

Organization of the paper

As explained above, the next two sections are devoted to positive results. In Section 4 we work out some bounds on the eigenvalues of random matrices, for later use in our reductions from dense subgraph problems to the approximation of RIP parameters. We rely mainly on the classical work of Füredi and Komlós [15] as well as on a more recent concentration inequality due to Alon, Krivelevich and Vu [4]. In Section 5 we present our two hypotheses on the intractability of dense subgraph problems, and we show that they are consistent with what we know from the spectral algorithm. In Section 7 we use these hypotheses to show that approximating RIP parameters is hard even for square matrices. In Section 8 we derive similar results for matrices of "strictly rectangular" format (which is the case of interest in compressed sensing). We proceed by reduction from the square case. Interestingly, this last reduction relies on the known constructions (deterministic [6, 7] and probabilistic [21]) of matrices with good RIP parameters mentioned earlier in the introduction. We therefore turn these positive results into negative results. The table at the end of Section 8 gives a summary of our hardness results. Section 6 is devoted to our skewed spectral algorithm.

This part of the paper can be read independently from the RIP results as it deals with a purely graph-theoretic problem. The analysis of the skewed algorithm is based on a new spectral norm estimate for a certain class of random matrices.

2 Increasing the order by decreasing the RIP parameter

As explained at the beginning of [6, 7], certain (suboptimal) constructions are based on the construction of systems of unit vectors $(u_1, \dots, u_N) \in \mathbb{C}^n$ with small coherence. The coherence parameter μ is defined as $\max_{i \neq j} |\langle u_i, u_j \rangle|$. Indeed, we have the following proposition.

Proposition 1 *Assume that the column vectors $u_1, \dots, u_N$ of Φ are of norm 1 and coherence μ. Then Φ satisfies the RIP of order k with parameter $\delta = (k-1)\mu$.*

We reproduce the proof from [6, 7] since if fits in one line: for any k-sparse vector x,

$$|||\Phi x||^2 - ||x||^2| \leq 2 \sum_{i<j} |x_i x_j \langle u_i, u_j \rangle| \leq \mu((\sum_i |x_i|)^2 - ||x||^2) \leq (k-1)\mu ||x||^2.$$

We now give a result, which (as we shall see) generalizes Proposition 1.

Theorem 1 *Assume that Φ has unit column vectors and satisfies the RIP of order m with parameter ϵ. For $k \geq m$, Φ also satisfies the RIP of order k with parameter $\delta = \epsilon(k-1)/(m-1)$.*

Proof. Let $u_1, \dots, u_N$ be the column vectors of Φ. Let x be a k-sparse vector, and write $x = \sum_{i \in T} x_i u_i$ where T is a subset of $\{1, \dots, N\}$ of size k. Since $||\Phi x||^2 = ||x||^2 + 2 \sum_{i<j} x_i x_j \langle u_i, u_j \rangle$, to check the RIP of order k we need to show that

$$|\sum_{i<j} x_i x_j \langle u_i, u_j \rangle| \leq \delta ||x||^2/2, \tag{1}$$

where $\delta = \epsilon(k-1)/(m-1)$. To estimate the left hand side, we compare it to the sum of the similar quantity taken over all subsets of size m of T, namely:

$$|\sum_{|S|=m} \sum_{i,j \in S, i<j} x_i x_j \langle u_i, u_j \rangle|. \tag{2}$$

Since each pair (i,j) appears in exactly $\binom{k-2}{m-2}$ subsets of size m, this sum is equal to $\binom{k-2}{m-2}$ times the left-hand side of (1). But we can also estimate (2) using the RIP of order m. For each subset S of size m, we have

$$|\sum_{i,j \in S, i<j} x_i x_j \langle u_i, u_j \rangle| \leq \epsilon \sum_{i \in S} x_i^2/2.$$

This follows from (1), replacing δ by ϵ (the RIP parameter of order m). Since each term x_i^2 will appear exactly in $\binom{k-1}{m-1}$ subsets, we obtain $\epsilon\binom{k-1}{m-1}||x||^2/2$ as an upper bound for (2). We conclude that the left-hand side of (1) is bounded by $\frac{\epsilon}{2}\binom{k-1}{m-1}||x||^2/\binom{k-2}{m-2} = \epsilon\frac{k-1}{m-1}||x||^2/2$. □

We claim that Proposition 1 is the case $m = 2$ of Theorem 1. This follows from the following observation.

Remark 1 *For a matrix Φ with unit column vectors, the coherence parameter μ is equal to the RIP parameter of order 2.*

Proof. Let δ be the RIP parameter of order 2. We have $\delta \leq \mu$ by Proposition 1. It remains to show that $\delta \geq \mu$. Consider therefore two column vectors u_i and u_j with $|\langle u_i, u_j \rangle| = \mu$. Let $x = u_i + u_j$. We have $||x||^2 = 2$ and $||\Phi x||^2 = 2 \pm 2\mu$, so that $\delta \geq \mu$ indeed. $\square$

3 A Matrix Certification Algorithm

The naive algorithm for computing the RIP parameter of order k will involve the enumeration of the $\binom{N}{k}$ submatrices of Φ made up of k column vectors of Φ. For each $T \subseteq \{1, \ldots, N\}$ of size k let us denote by Φ_T the corresponding $n \times k$ matrix. We need to compute (or upper bound) $\delta = \max_T \delta_T$, where

$$\delta_T = \sup_{x \in \mathbb{C}^k} | \; ||\Phi_T x||^2 / ||x||^2 - 1 \; |.$$

For each T, δ_T can be computed efficiently by linear algebra. For instance, δ_T is the spectral radius of the self-adjoint matrix $\Phi_T^* \Phi_T - \mathbf{I}_k$. The cost of the computation is therefore dominated by the combinatorial factor $\binom{N}{k}$ due to the enumeration of all subsets of size k.

Here we analyze what the naive algorithm can gain from Theorem 1. We therefore consider the following *lazy algorithm.* The correctness of the algorithm follows immediately from Theorem 1.

Algorithm 1

1: **procedure** LAZY(Φ, m, δ)
2: **Input:** a $n \times N$ matrix Φ with unit column vectors, an integer $m \leq n$, and a parameter $\delta \in]0, 1[$.
3: Compute as explained above the RIP parameter of order m. Call it ϵ.
4: **Output:** Certify Φ as a RIP matrix of order k with parameter δ, for all $k \geq m$ such that $\epsilon(k-1)/(m-1) \leq \delta$.
5: **end procedure**

We now analyze its behavior on random matrices, which are in many cases known to satisfy the RIP with high probability. Consider for instance the case of a matrix whose entries are independent symmetric Bernouilli random variables.

Theorem 2 *Let A be a $n \times N$ matrix whose entries are independent symmetric Bernouilli random variables and assume that $n \geq C\epsilon^{-2} m \log(eN/m)$. With probability at least $1 - 2\exp(-c\epsilon^2 n)$, the normalized matrix $\Phi = \frac{1}{\sqrt{n}} A$ satisfies the the RIP of order m with parameter ϵ. Here C and c are absolute constants.*

In fact the same theorem holds for a very large class of random matrix models, namely, subgaussian matrices with either independent rows or independent columns ([21], Theorem 64).

Proposition 2 *Let A be a random matrix as in Theorem 2, and $\delta \in]0, 1[$. With probability at least $1 - 2(eN/m)^{-cCm}$, the lazy algorithm presented above will certify that A satisfies the RIP of order k with parameter δ for all k such that:*

$$k \leq \delta \sqrt{\frac{mn}{c \log(eN/m)}}.$$

Here c and C are the absolute constants from Theorem 2.

Proof. All parameters being fixed we take ϵ as small as allowed by Theorem 2, so that $n\epsilon^2 = Cm\log(eN/m)$. This yields the announced probability estimate, and the upper bound on k is $\delta m/\epsilon$. $\square$

To compare the lazy algorithm to the naive algorithm, set for instance $m = \sqrt{n}$. In applications to compressed sensing one can set δ to a small constant value (any $\delta < \sqrt{2} - 1$ will do). Thus, disregarding constant and logarithmic factors, with high probability the lazy algorithm will certify the RIP property for k of order roughly $n^{3/4}$. This is achieved by enumerating $\binom{N}{n^{1/2}}$ subspaces, whereas the naive algorithm would enumerate roughly $\binom{N}{n^{3/4}}$ subspaces.

Another choice of parameters in Proposition 2 shows that one can beat the $\sqrt{n}$ bound by a logarithmic factor with a quasi-polynomial time algorithm. For instance:

Corollary 1 *If we set $m = (\log N)^3$, the lazy algorithm runs in time $2^{O(\log^4 N)}$ and, with probability at least $1 - 2^{-\Omega((\log N)^4)}$ certifies that A satisfies the RIP of order k with parameter δ for all $k \leq K\delta \log N\sqrt{n}$, where K is an absolute constant.*

4 Eigenvalues of Random Symmetric Matrices

Proposition 4 is the main probabilistic inequality that we derive in this section. It shows that square matrices obtained by Cholesky decomposition from a certain class of random matrices have good RIP parameters with high probability. This result is then used in Section 7 to give a reduction from dense subgraph problems to the approximation of RIP parameters.

4.1 Model A

Consider the following random matrix model: A is a symmetric $k \times k$ matrix with $a_{ii} = 0$, and for $i < j$ the a_{ij} are independent symmetric Bernouilli random variables.

Let $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \lambda_k(A)$ be the eigenvalues of A. Let m_s be the median of $\lambda_s(A)$. From the main result of [4] (bottom of p. 263) we have for $t \geq 0$ the inequality:

$$\Pr[\lambda_s(A) - m_s \geq t] \leq 2e^{-t^2/32s^2}.$$

From Füredi and Komlós ([15], Theorem 2) we know that $m_1 \leq 3\sigma\sqrt{k}$ for k large enough, where $\sigma = 1$ is the standard deviation of the a_{ij} in the case $i < j$. Therefore we have

$$\Pr[\lambda_1(A) \geq 3\sqrt{k} + t] \leq 2e^{-t^2/32}.$$

Since $\lambda_k(A) = -\lambda_1(-A)$ and $-A$ has same distribution as A, we also have

$$\Pr[\lambda_k(A) \leq -3\sqrt{k} - t] \leq 2e^{-t^2/32}$$

(one could also apply directly the bound on $\lambda_k(A)$ for the more general model considered in [4]). As a result:

Proposition 3 *There is an integer k_0 such that for all $k \geq k_0$ and for all $t \geq 0$ we have:*

$$\Pr[\max_i |\lambda_i(A)| \geq 3\sqrt{k} + t] \leq 4e^{-t^2/32}.$$

Remark 2 *The constant 3 in Proposition 3 can be replaced by any constant bigger than 2 (see Theorem 2 in [15]).*

4.2 Model B

Next we consider the model where B is a symmetric $k \times k$ matrix satisfying the following condition: $b_{ii} = 1$, and $b_{ij} = c \cdot a_{ij}/\sqrt{n}$ for $i < j$, where the a_{ij} are independent symmetric Bernouilli random variables. Here $c > 0$ is a fixed constant, and n is an additional parameter which should be thought of as going to infinity with k.

Corollary 2 *Assume that $k \geq k_0$ and that $\delta\sqrt{n} \geq 3c\sqrt{k}$. Then the eigenvalues of B all lie in the interval $[1-\delta, 1+\delta]$ with probability at least*

$$1 - 4\exp[-(\frac{\delta\sqrt{n}}{c} - 3\sqrt{k})^2/32].$$

Proof. We have $B = \mathbf{I}_k + cA/\sqrt{n}$, where A follows the model of Proposition 3. The result therefore follows from that proposition by choosing t so that $c(3\sqrt{k}+t)/\sqrt{n} = \delta$, i.e., $t = \delta\sqrt{n}/c - 3\sqrt{k}$. □

In the next corollary we look at the case $n = k$ of this model.

Corollary 3 *Assume that $n \geq k_0$ and $3c < 1$. Then B is positive semi-definite with probability at least*

$$1 - 4\exp[-(1/c - 3)^2 n/32].$$

Proof. Set $n = k$ and $\delta = 1$ in Corollary 2. □

In the last result of this subsection we consider again the model $B = \mathbf{I}_n + cA/\sqrt{n}$. Given a $n \times n$ matrix M and two subsets $S, T \subseteq \{1, \ldots, n\}$ of size k, let us denote by $M_{S,T}$ the $k \times k$ sub-matrix made up of all entries of M of row number in S and column number in T.

Corollary 4 *Consider the random matrix $B = \mathbf{I}_n + cA/\sqrt{n}$ where A is drawn from the uniform distribution on the set $n \times n$ symmetric matrices with null diagonal entries and ± 1 off-diagonal entries.*

If $n \geq k \geq k_0$, then with probability at least

$$1 - 4\exp\left[k\ln(ne/k) - (\frac{\delta\sqrt{n}}{c} - 3\sqrt{k})^2/32\right]$$

the submatrices $B_{S,S}$ have all their eigenvalues in the interval $[1-\delta, 1+\delta]$ for all subsets $S \subseteq \{1, \ldots, n\}$ of size k.

Proof. By Corollary 2, for each fixed S matrix $B_{S,S}$ has an eigenvalue outside of the interval $[1-\delta, 1+\delta]$ with probability at most $4\exp[-(\frac{\delta\sqrt{n}}{c} - 3\sqrt{k})^2/32]$. The result follows by taking a union bound over the $\binom{n}{k} \leq (ne/k)^k$ subsets of size k. □

4.3 Model C

In Corollaries 3 and 4 we considered the following random model for B: set $B = \mathbf{I}_n + cA/\sqrt{n}$, where A is chosen from the uniform distribution on the set S_n of all symmetric matrices with null diagonal entries and ± 1 off-diagonal entries. If B is positive semi-definite, we can find by Cholesky decomposition a $n \times n$ matrix C such that $C^T C = B$. If B is not positive semi-definite, we set $C = 0$. This is the random model for C that we study in this subsection.

Proposition 4 *Assume that $n \geq k \geq k_0$ and that $3c < \min(1, \delta\sqrt{n}/\sqrt{k})$. With probability at least*

$$1 - 4\exp\left[k\ln(ne/k) - (\frac{\delta\sqrt{n}}{c} - 3\sqrt{k})^2/32\right] - 4\exp[-(1/c-3)^2 n/32],$$

C satisfies the RIP of order k with parameter δ.

Proof. If $B = \mathbf{I}_n + cA/\sqrt{n}$ is not positive semi-definite then $C = 0$ and this matrix obviously does not satisfy the RIP. By Corollary 3, B can fail to be positive semi-definite with probability at most $4\exp[-(1/c-3)^2 n/32]$.

If B is positive semi-definite then $C^T C = B$. Using the notation of Corollary 4, matrix C satisfies the RIP of order k with parameter δ if for all subsets S of size k, the eigenvalues of the $k \times k$ matrices $(C^T C)_{S,S}$ all lie in the interval $[1-\delta, 1+\delta]$. Since $C^T C = B$, by Corollary 4 this can happen with probability at most $4\exp[k\ln(ne/k) - (\frac{\delta\sqrt{n}}{c} - 3\sqrt{k})^2/32]$. □

5 Dense Subgraph Problems

In this section we formulate precisely our two hypotheses on the intractability of dense subgraph problems. We show that they are compatible with what follows from the spectral algorithm (see Theorem 3 below). Spectral methods have become classical in graph theory, see for instance [3] for a nontrivial application to the detection of large hidden cliques in random graphs.

We work with the standard $G(n, 1/2)$ random graph model: each of the $n(n-1)/2$ potential edges is chosen with probability $1/2$, and these choices are made independently.

Definition 1 *We say that a graph H on k vertices has density λ if it has at least $\lambda k(k-1)/2$ edges. For $\epsilon \geq 0$, we say that G has excess ϵ if it has density $1/2 + \epsilon$.*

Lemma 1 *In a random graph on n vertices, the probability of existence of a k-subgraph with excess ϵ is upper bounded by*

$$\exp[k\ln(ne/k) - \epsilon^2 k(k-1)].$$

Proof. Fix any subgraph H on k vertices. The number of edges in H is a sum of $k(k-1)/2$ independent Bernouilli random variables. Therefore, by Hoeffding's inequality the probability for H to have excess ϵ can be upper bounded by $\exp[-\epsilon^2 k(k-1)]$. The result follows by taking a union bound over the $\binom{n}{k} \leq (ne/k)^k$ subgraphs of size k. □

This probability goes rapidly to 0 once ϵ becomes significantly larger than $1/\sqrt{k}$. For instance:

Corollary 5 *Fix any constant $c > 0$. For $\epsilon \geq (1+c)\sqrt{\ln(ne/k)/k}$, the probability that a random graph on n vertices has a k-subgraph with excess ϵ goes to 0 as $k, n \to +\infty$.*

When ϵ is in the range of Corollary 5, it makes sense to consider algorithm which will certify that most graphs on n vertices do not have any k-subgraph with excess ϵ (i.e., a certified graph should not have any k-subgraph with excess ϵ, and most graphs should be certified).

As we shall now see (in Theorem 3), the spectral method provides an efficient certification algorithm if $\epsilon k/\sqrt{n}$ is sufficiently large.

Lemma 2 *Let A be the signed adjacency matrix of a graph G. If G has a k-subgraph H with excess ϵ then there is a unit vector $x \in \mathbb{R}^n$ supported by k basis vectors such that $x^T Ax \geq 2\epsilon(k-1)$.*

Proof. Assume without loss of generality that $V(G) = \{1, \ldots, n\}$ and $V(H) = \{1, \ldots, k\}$. We have $a_{ii} = 0$ and for $i \neq j$, $a_{ij} = 1$ if $ij \in E$; $a_{ij} = -1$ if $ij \notin E$. Let $x \in \mathbb{R}^n$ be the unit vector uniformly supported by the first k basis vectors: $x_i = 1/\sqrt{k}$ for $i \leq k$, $x_i = 0$ for $i > k$. Then $x^T Ax = 2\sum_{1 \leq i < j \leq k} a_{ij}/k \geq 2\epsilon(k-1)$. □

Corollary 6 *Under the same hypothesis as in Lemma 2, the largest eigenvalue of A satisfies $\lambda_1(A) \geq 2\epsilon(k-1)$.*

Proof. This follows directly from the lemma since for a symmetric matrix, $\lambda_1(A) = \sup_{||x||=1} x^T Ax$. □

This leads to the following polynomial-time certification algorithm.

Algorithm 2

1: **procedure** SPECTRAL(G)
2: **Input:** a graph G on n vertices.
3: Compute $\lambda_1(A)$, where A is the signed adjacency matrix of G.
4: **Output:** Certify that G does not contain any k-subgraph with excess ϵ for all k, ϵ such that $2\epsilon(k-1) > \lambda_1(A)$.
5: **end procedure**

Theorem 3 *The spectral algorithm presented above certifies that most graphs on n vertices do not contain any k-subgraph with excess ϵ for all k, ϵ such that*

$$2\epsilon(k-1) \geq 3\sqrt{n}.$$

Proof. The correctness of the algorithm follows from Corollary 6. By Theorem 2 in [15] most graphs satisfy $\lambda_1(A) < 3\sqrt{n}$, so most graphs are certified. □

Remark 3 *The constant 3 in this theorem can be replaced by any constant larger than 2 (see also Remark 2).*

In the case $\epsilon = 1/2$, the spectral algorithm of Theorem 3 certifies that most graphs on n vertices do not contain any k-clique for $k \geq 1 + 3\sqrt{n}$.

We put forward the following hypotheses.

Hypothesis 1 *Fix two arbitrary rational constants $\alpha, \beta \in]0,1[$ such that $\alpha - \beta < 1/2$. Set $\epsilon(n) = 1/n^\beta$ and $k(n) = n^\alpha$ (we only consider values of n such that $k(n) \in \mathbb{N}$).*

No polynomial time algorithm can certify that most graphs on n vertices do not have any $k(n)$-subgraph with excess $\epsilon(n)$.

Hypothesis 2 *There is a constant $\kappa \in]0,1[$ such that the following holds: fix two arbitrary constants $\alpha \in]0,1[$ and $\gamma > 0$ with $\alpha \leq \kappa$ and $\alpha\gamma \leq \kappa$. Set $\epsilon(n) = \gamma/\sqrt{n}$ and $k(n) = \alpha n$ (again we only consider values of n such that $k(n) \in \mathbb{N}$). Then no polynomial time algorithm can certify that most graphs on n vertices do not have any $k(n)$-subgraph with excess $\epsilon(n)$.*

The constraints on the parameters α, β, γ in these two hypotheses are meant to ensure consistency with what we know from the spectral algorithm: in Hypothesis 1 we have $\epsilon k = n^{\alpha-\beta} < \sqrt{n}$. In Hypothesis 2 we have $\epsilon k = \alpha\gamma/\sqrt{n} \leq \kappa\sqrt{n} < \sqrt{n}$. By Remark 3, if we allowed a value $\kappa > 1$ in Hypothesis 2 then this hypothesis would become provably incorrect. Hypothesis 2 essentially means that the spectral algorithm can be improved at most by a constant factor. As we will see later, both hypotheses are also consistent with the skewed algorithm of Section 6.

Of course, our two hypotheses are meaningful only for values of the parameters for which it is actually the case that most graphs do not have a k-subgraph with excess ϵ (otherwise, they are vacuously true). By Corollary 5, for the first hypothesis it is enough to have $\beta < \alpha/2$. Therefore we can first pick any $\beta \in]0, 1/2[$ and then any $\alpha \in]2\beta, \beta + 1/2[$ (the upper bound $\beta + 1/2$ enforces the constraint $\alpha - \beta < 1/2$ in Hypothesis 1).

Let us now work out the corresponding constraints for Hypothesis 2. Take for instance $c = 1$ in Corollary 5. Plugging in $k = \alpha n$ and $\epsilon = \gamma/\sqrt{n}$ yields the condition

$$\gamma \geq 2\sqrt{\frac{\ln(e/\alpha)}{\alpha}}.$$

But we also have the constraint $\gamma \leq \kappa/\alpha$ from Hypothesis 2. So we should have

$$\kappa/\alpha \geq 2\sqrt{\frac{\ln(e/\alpha)}{\alpha}}.$$

This constraint is satisfied by all sufficiently small α. Hence we can first pick any small enough α, and then any γ in the interval $[2\sqrt{\ln(e/\alpha)/\alpha}, \kappa/\alpha]$.

6 A Skewed Spectral Algorithm

Let G be an undirected graph on n vertices and A its signed adjacency matrix. We define the skewed adjacency matrix of G by the formula

$$\widehat{A} = \frac{A}{2} + \frac{a}{\sqrt{n}}\mathbf{J}, \tag{3}$$

where $\mathbf{J}$ is the $n \times n$ all 1's matrix, and $a \geq 0$ a parameter to be tuned later. We call the matrix $\widehat{A}$ skewed because its entries are random variables with mean slightly above zero.

Lemma 3 *If G has a k-subgraph H with excess ϵ then there is a unit vector $x \in \mathbb{R}^n$ supported by k basis vectors such that $x^T \widehat{A} x \geq \epsilon(k-1) + ka/\sqrt{n}$.*

Proof. Assume without loss of generality that $V(G) = \{1, \ldots, n\}$ and $V(H) = \{1, \ldots, k\}$. Let $x \in \mathbb{R}^n$ be the unit vector uniformly supported by the first k basis vectors: $x_i = 1/\sqrt{k}$ for $i \leq k$, $x_i = 0$ for $i > k$. We have seen that $x^T A x \geq 2\epsilon(k-1)$. The conclusion follows since $x^T \mathbf{J} x = k$. □

This simple lemma forms the basis for the following *skewed algorithm*. In the case $a = 0$, this is just the standard spectral algorithm. Since $\widehat{A}$ is symmetric, $\lambda_1(\widehat{A}) = \sup_{\|x\|=1} x^T \widehat{A} x$. The correctness of the skewed algorithm therefore follows from Lemma 3. In the remainder of this section, we analyze the behavior of this algorithm on random graphs.

Algorithm 3

1: **procedure** SKEWED SPECTRAL(G, a)
2: **Input:** a graph G on n vertices.
3: Compute $\lambda_1(\widehat{A})$, where as above $\widehat{A}$ is the skewed adjacency matrix of G.
4: **Output:** Certify that G does not contain any k-subgraph with excess ϵ for all k, ϵ such that $\epsilon(k-1) + ka/\sqrt{n} > \lambda_1(\widehat{A})$.
5: **end procedure**

Theorem 4 *Let G be a random graph on n vertices and $\widehat{A}$ its skewed adjacency matrix as defined by (3). The expectation of the operator norm $\left\|\widehat{A}\right\| = \sup_{\|x\|=1} \left\|\widehat{A}x\right\|$ satisfies the inequality*

$$\mathbb{E}\left\|\widehat{A}\right\| \leq \sqrt{n}\sqrt{a^2 + 5/4 + o(1)}.$$

Moreover, for all $\epsilon \geq 0$ we have the concentration inequality

$$\Pr\left[\left\|\widehat{A}\right\| \geq \mathbb{E}\left\|\widehat{A}\right\| + \epsilon\right] \leq \exp(-\epsilon^2/8).$$

We defer the proof of the above theorem to the appendix. Compare this bound on $\mathbb{E}\left\|\widehat{A}\right\|$ with the bound $(a+1+o(1))\sqrt{n}$ that follows from applying the triangle inequality on $\widehat{A}$. A direct calculation shows that the bound in Theorem 4 is better for $a > 1/8$.

For any matrix, the largest eigenvalue is upper bounded by the operator norm. We therefore have the following corollary.

Corollary 7 *For most graphs G on n vertices, the skewed adjacency matrix of G satisfies the inequality*

$$\lambda_1(\widehat{A}) \leq \sqrt{n}\sqrt{a^2 + 5/4 + o(1)}.$$

Now, let k and ϵ be two functions of n. From Corollary 7 and the correctness of the skewed algorithm it follows that whenever the inequality

$$\liminf_{n\to+\infty} \frac{1}{\sqrt{n}}\left[\epsilon(n).(k(n)-1) + \frac{ak(n)}{\sqrt{n}}\right] > \sqrt{a^2 + 5/4} \tag{4}$$

is satisfied, the skewed algorithm can certify that most graphs on n vertices do not have any $k(n)$-subgraph with excess $\epsilon(n)$. In particular, we have obtained the following result.

Theorem 5 *Set $\epsilon(n) = \gamma/\sqrt{n}$ and $k(n) = \alpha n$, where $\alpha \in]0,1[$ and $\gamma > 0$ are two constants. If $\alpha\gamma + \alpha a > \sqrt{a^2 + 5/4}$ then the skewed algorithm can certify that most graphs on n vertices do not have any $k(n)$-subgraph with excess $\epsilon(n)$.*

For instance, taking $a = 1$ yields the following.

Corollary 8 *If $\alpha\gamma + \alpha > 3/2$ then the skewed algorithm can certify that most graphs on n vertices do not have any αn-subgraph with excess $\gamma/\sqrt{n}$.*

This is clearly an improvement over the standard spectral algorithm, since that algorithm is unable to certify that most graphs on n vertices do not have any αn-subgraph with excess $\gamma/\sqrt{n}$ when $\alpha\gamma < 1$. But there are values $\alpha \in]0,1[$ and $\gamma > 0$ for which $\alpha\gamma < 1$ and the condition of Corollary 8 is satisfied (take for instance $\alpha = 3/4$ and $\gamma = 5/4$).

More generally, given values of α and γ we can ask when there is a value of a such that the skewed algorithm can certify that most graphs on n vertices do not have any αn-subgraph with excess $\gamma/\sqrt{n}$.

Corollary 9 *If the condition $4\alpha^2\gamma^2 + 5\alpha^2 > 5$ is satisfied, the skewed algorithm can certify that most graphs on n vertices do not have any αn-subgraph with excess $\gamma/\sqrt{n}$.*

Proof. In view of Theorem 5, we are looking for a value of a such that

$$f(a) = a^2 + 5/4 - (\alpha\gamma + \alpha a)^2 < 0.$$

There is such an a if $\Delta > 0$, where Δ is the discriminant of the quadratic polynomial $f(a)$. But

$$\Delta = 4\alpha^4\gamma^2 - 4(1-\alpha^2)(5/4 - \alpha^2\gamma^2) = -5 + 4\alpha^2\gamma^2 + 5\alpha^2.$$

□

Note that this analysis of the skewed algorithm is suboptimal: in the case $a = 0$ we can certify with the standard spectral algorithm that most graphs on n vertices do not have any αn-subgraph with excess $\gamma/\sqrt{n}$ as soon as $\alpha\gamma > 1$. But this fact cannot be derived from Corollary 9. In order to improve this corollary, one would need to improve the upper bound on $\mathbb{E}\left\|\widehat{A}\right\|$ given by Theorem 4.

It follows from Corollary 9 that the constant κ in Hypothesis 2 must satisfy the constraint $9\kappa^2 \leq 5$, i.e., $\kappa \leq \sqrt{5}/3 \simeq 0.745$.

Finally, let us consider the case where (as in Hypothesis 1) the two functions $k(n)$ and $\epsilon(n)$ are of the form $k(n) = n^\alpha$ and $\epsilon(n) = 1/n^\beta$ for two constants $\alpha, \beta \in]0,1[$. With those settings, the $\liminf$ in (4) is equal to 0 if $\alpha - \beta < 1/2$, to 1 if $\alpha - \beta = 1/2$ and to $+\infty$ if $\alpha - \beta > 1/2$. Hence we do not obtain any improvement over the standard spectral algorithm (both algorithms succeed only in the case $\alpha - \beta > 1/2$).

7 Dense Subgraphs and the Restricted Isometry Property

In this section we show (in Theorems 6 and 7) that RIP parameters are hard to approximate even for square matrices. We establish connections between dense subgraphs problems and the RIP thanks to a generic reduction which we call the *Cholesky reduction*. This reduction maps a graph G on n vertices to a $n \times n$ matrix $C(G)$. Let A be the signed adjacency matrix of G: we have $a_{ii} = 0$ and for $i \neq j$, $a_{ij} = 1$ if $ij \in E$; $a_{ij} = -1$ if $ij \notin E$. We construct $C = C(G)$ from A using the procedure described in Section 4.3. That is, we first compute $B = \mathbf{I}_n + cA/\sqrt{n}$ (suitable choices for the parameter c will be discussed later). If B is not positive semi-definite, we set $C = 0$. Otherwise, we find by Cholesky decomposition a matrix C such that $C^TC = B$.

For suitable values of k, $C(G)$ satisfies the RIP of order k for most graphs G. This was made precise in Proposition 4. On the other hand, we shall see that if G has a k-subgraph that is too dense then $C(G)$ does not satisfy the RIP of order k.

Proposition 5 *Let G be a graph on n vertices and $C(G)$ its image by the Cholesky reduction. If G has a k-subgraph with excess ϵ and*

$$\delta < 2c\epsilon(k-1)/\sqrt{n}$$

then $C(G)$ does not satisfy the RIP of order k with parameter δ.

Proof. Let A be the signed adjacency matrix of G and $B = \mathbf{I}_n + cA/\sqrt{n}$. If B is not semi-definite positive, $C(G) = 0$ does not satisfy the RIP. Otherwise $C^TC = B$. Let x be the vector of Lemma 2. We have $||Cx||^2 = x^TC^TCx = x^TBx = 1 + cx^TAx/\sqrt{n} > 1 + \delta$. □

7.1 Hardness from Hypothesis 1

Lemma 4 *Set $k = n^\alpha$ and $\epsilon = 1/n^\beta$ where (as suggested in Section 5) α, β are two constants satisfying the constraints $\beta \in]0, 1/2[$ and $\alpha \in]2\beta, \beta + 1/2[$. Set also $\delta = c'\epsilon k/\sqrt{n}$ where $c' > 0$ is another constant. Then for most graphs G on n vertices the matrix $C(G)$ satisfies the RIP of order k with parameter δ.*

Proof. Fix a constant $c < 1/3$. We can apply Proposition 4 since the hypothesis $3c < \min(1, \delta\sqrt{n}/\sqrt{k})$ is satisfied for n large enough (note that $\delta\sqrt{n}/\sqrt{k} = c'\epsilon\sqrt{k} = c'n^{\alpha/2-\beta}$, and the exponent $\alpha/2 - \beta$ is positive).

In Proposition 4, the probability that $C = C(G)$ does *not* satisfy the RIP is bounded by a sum of two terms. The second one, $4\exp[-(1/c - 3)^2 n/32]$, clearly goes to 0 as $n \to +\infty$. To obtain the same property for the first term we consider the argument:

$$k\ln(ne/k) - (\frac{\delta\sqrt{n}}{c} - 3\sqrt{k})^2/32. \tag{5}$$

Note that $\delta\sqrt{n}/c = (c'/c)\epsilon k = (c'/c)n^{\alpha-\beta}$, so that this term dominates the term $3\sqrt{k} = 3n^{\alpha/2}$ since $\alpha - \beta > \alpha/2$. Moreover $(\delta\sqrt{n}/c)^2 = (c'/c)^2 n^{2(\alpha-\beta)}$ dominates the term $k\ln(ne/k)$ since $k = n^\alpha$ and $2(\alpha - \beta) > \alpha$. Therefore (5) is equivalent to $-(c'/c)^2 n^{2(\alpha-\beta)}/32$, and its exponential goes to 0 as n goes to infinity. Hence the same is true for the sum of the two error terms. □

Theorem 6 *Under Hypothesis 1, no polynomial-time algorithm can approximate the RIP parameters of square matrices within any constant factor.*

More precisely, fix any constants $\lambda > 1$ and $\delta_0 < 1/\lambda$. Under Hypothesis 1, no polynomial time algorithm can distinguish between matrices which satisfy the RIP of order k with parameter $\delta \leq \delta_0$ and matrices which do not satisfy the RIP of order k with parameter $\lambda\delta$.

In this theorem a distinguishing algorithm must accept all inputs which satisfy the RIP of order k with parameter $\delta \leq \delta_0$, and reject all inputs which do not satisfy the RIP of order k with parameter $\lambda\delta$. Its behavior on other inputs is not specified.

Proof of Theorem 6. The second part of the theorem clearly implies the first: if we have an algorithm which approximates RIP parameters within some factor $\gamma > 1$ then we have a distinguishing algorithm with $\lambda = \gamma^2$.

Let us therefore assume that we have a distinguishing algorithm $\mathcal{A}$ with the associated constants λ and δ_0. Fix a positive constant $c < 1/3$ and two constants α and β satisfying the constraints

of Lemma 4, for instance $\beta = 1/3$ and $\alpha = 3/4$. Set as usual $k = n^\alpha$ and $\epsilon = 1/n^\beta$. We will use $\mathcal{A}$ to certify that most graphs on n vertices do not have any k-subgraph with excess ϵ, thereby contradicting Hypothesis 1.

On the one hand, by Lemma 4 for most graphs G the matrix $C(G)$ satisfy the RIP of order k with parameter $\delta = c'\epsilon k/\sqrt{n}$ (we will see quite soon that c' can be any sufficiently small constant). In this case $\delta = c'\epsilon k/\sqrt{n} = c'n^{\alpha-\beta-1/2}$ goes to 0 as $n \to +\infty$, so for n large enough we have $\delta \leq \delta_0$ and we can use the hypothesis on our distinguishing algorithm.

On the other hand, if G has a k-subgraph with excess ϵ then by Proposition 5 the RIP parameter of $C(G)$ is not smaller than $2c\epsilon(k-1)/\sqrt{n}$. This is more than $\lambda\delta$ if $\lambda c' < 2c$ and n is large enough. Therefore given λ and c we choose a constant $c' < 2c/\lambda$, and by running $\mathcal{A}$ on $C(G)$ we can certify that most graphs on n vertices do not have any k-subgraph with excess ϵ (we certify G if $\mathcal{A}$ accepts $C(G)$). $\square$

Note that the above hardness result has been established for *very small* RIP parameters: the matrices in the proof have RIP parameters of order $\epsilon k/\sqrt{n} = n^{\alpha-\beta-1/2}$ and the exponent $\alpha-\beta-1/2$ is negative. This may be viewed as a weakness of the result since applications to compressed sensing only require a constant δ (as pointed out in the introduction, any $\delta < \sqrt{2}-1$ will do). We partially overcome this weakness in Section 7.2, but for this we need to replace Hypothesis 1 by Hypothesis 2.

Note also that the proof of Theorem 6 yields more information than contained in the statement of the theorem. For instance, with the settings $\alpha = 3/4$ and $\beta = 1/3$ we see that the gap problem in Theorem 6 remains hard even for $k = n^{3/4}$ and $\delta = c'/n^{1/12}$. We have included this additional information in the table at the end of the paper, for Theorem 6 as well as for our other hardness results.

7.2 Hardness from Hypothesis 2

Lemma 5 *Set $k = \alpha n$ and $\epsilon = \gamma/\sqrt{n}$ where $\alpha \in]0,1[$ and $\gamma > 0$ are two constants. Fix a constant $c < 1/3$ and another constant c' such that $3c < c'\gamma\sqrt{\alpha}$ and*

$$\ln(e/\alpha) < (\frac{c'\gamma\sqrt{\alpha}}{c} - 3)^2/32. \tag{6}$$

Finally, set $\delta = c'\epsilon k/\sqrt{n} = c'\alpha\gamma$. Then for most graphs G on n vertices the matrix $C(G)$ satisfies the RIP of order k with parameter δ.

Proof. We can apply Proposition 4 since the hypothesis $3c < \min(1, \delta\sqrt{n}/\sqrt{k})$ is satisfied: $\delta\sqrt{n}/\sqrt{k} = \delta/\sqrt{\alpha} = c'\gamma\sqrt{\alpha} > 3c$ by choice of c'.

As in the proof of Lemma 4, the second error term in Proposition 4 clearly goes to 0 as $n \to +\infty$. We therefore turn our attention to the first error term, and in particular to the argument (5) of the exponential function.

In (5) the term $\delta\sqrt{n}/c - 3\sqrt{k}$ is equal to $(c'\alpha\gamma/c - 3\sqrt{\alpha})\sqrt{n}$ so the term $(\frac{\delta\sqrt{n}}{c} - 3\sqrt{k})^2/32$ is equal to $\alpha(c'\sqrt{\alpha}A/c - 3)^2 n/32$.

The positive term in (5) is $k\ln(ne/k) = \alpha n \ln(e/\alpha)$. Overall, (5) is equal to $-Kn$ for some constant K, and K is positive by (6). $\square$

Theorem 7 *Fix any constant $\delta_0 < \kappa/2$, where κ is the constant from Hypothesis 2. Assuming this hypothesis, no polynomial time algorithm can distinguish between matrices which satisfy the RIP of order k with parameter δ_0 and matrices which do not satisfy the RIP of order k with parameter $\kappa/2$.*

Proof. Let us assume the contrary, and let $\mathcal{A}$ be the distinguishing algorithm. Set $k = \alpha n$ and $\epsilon = \gamma/\sqrt{n}$ where $\gamma = \kappa/\alpha$ and the constant $\alpha \in]0, \kappa[$ is small enough (we will see at the end of the proof how small is small enough). We will use $\mathcal{A}$ to certify that most graphs on n vertices do not have any k-subgraph with excess ϵ, thereby contradicting Hypothesis 2.

Consider first the case of a graph G having a k-subgraph with excess ϵ. Let us fix a constant $c < 1/3$ such that $2c\kappa > \kappa/2$. Then by Proposition 5 the matrix $C(G)$ does not satisfy the RIP of order k with parameter $\kappa/2$.

On the other hand, if we can find a constant c' which satisfies the hypotheses of Lemma 5 and such that $\delta = c'\alpha\gamma = \kappa c' < \delta_0$ then $C(G)$ will satisfy the RIP of order k with parameter δ_0 for most graph G. We will therefore be able to use $\mathcal{A}$ to certify that most graphs on n vertices do not have any k-subgraph with excess ϵ.

It just remains to explain how to choose c'. We will simply pick any c' such that $\kappa c' < \delta_0$ and will then check that the hypotheses of Lemma 5 are satisfied if α is small enough. First we have to check the condition $3c < c'\gamma\sqrt{\alpha}$. The right hand side is equal to $\kappa c'/\sqrt{\alpha}$ and will exceed the left-hand side if e.g. $\sqrt{\alpha} < c'\kappa$ (recall that $3c < 1$). The other condition to be checked is (6). For the same reason, it will be satisfied for small enough α: the right-hand side is equivalent to $(\kappa c'/c)^2/(32\alpha)$, which dominates the logarithmic function on the left-hand side. □

8 Hardness for Rectangular Matrices and Randomized Certification

In this section we show that the RIP parameters of rectangular matrices are hard to approximate. This is the case of interest in compressed sensing. In a sense this was already done in Section 7: we have shown that the special case of square matrices is already hard. Nevertheless, it is of interest to know that the problems remains hard for *strictly rectangular* matrices. This is what we do in this section. Proofs are essentially by reduction from the square case. We begin with a simple lemma.

Lemma 6 *Consider a matrix Φ with the following block structure:*

$$\Phi = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

This matrix satisfies the RIP of order k with parameter δ if and only if the same is true for both A and B.

Proof. For an input vector x with the corresponding block structure $x = (u\ v)$ we have $||x||^2 = ||u^2|| + ||v||^2$ and $||\Phi x||^2 = ||Au||^2 + ||Bv||^2$. Therefore, if Φ satisfies the RIP of order k with parameter δ then the same is true for A (take $v = 0$ and u k-sparse). The same argument applies also to B.

Conversely, assume that A and B satisfy the RIP of order k with parameter δ. Let $x = (u\ v)$ be a k-sparse vector. We have $||\Phi x||^2 - ||x||^2 = (||\Phi u||^2 - ||u||^2) + (||\Phi v||^2 - ||v||^2)$. Both u and v must be k-sparse, so the first term is bounded in absolute value by $\delta||u||^2$ and the second one by $\delta||v||^2$. The result follows since $||u||^2 + ||v||^2 = ||x||^2$. □

Theorem 8 *Fix any constants $\lambda > 1$ and $\delta_0 < 1/\lambda$. Under Hypothesis 1, no polynomial time algorithm can distinguish between matrices which satisfy the RIP of order k with parameter $\delta \leq \delta_0$ and matrices which do not satisfy the RIP of order k with parameter $\lambda\delta$.*

Moreover, polynomial-time distinction between these two cases remains impossible even for matrices of size $2n \times (n+N)$ *where* $N = n^{1+\epsilon_0}$. *Here* $\epsilon_0 \in]0, 1/2[$ *is an absolute constant (independent of* λ *and* δ_0*).*

The first part of this theorem is just Theorem 6. The second part is new. Its proof is a refinement of the proof of Theorem 6 (and refers to it).

Proof of Theorem 8. From a graph G on n vertices we construct the matrix

$$C'(G) = \begin{pmatrix} C(G) & 0 \\ 0 & B_n \end{pmatrix}$$

where $C(G)$ is as in the previous section and B_n is a matrix with good restricted isometry properties. Its role is to ensure the rectangular format that we need for $C'(G)$. Our specific choice for B_n is the matrix constructed in [6, 7]. It is of size $n \times N$ where $N = n^{1+\epsilon_0}$, and it satisfies the RIP of order $n^{\frac{1}{2}+\epsilon_0}$ with parameter $n^{-\epsilon_0}$. Moreover, B_n can be constructed deterministically in time polynomial in n. Note that $C'(G)$ is of size $2n \times (n+N)$ as required in the statement of Theorem 8.

Let us assume that we have a distinguishing algorithm $\mathcal{A}$ which works for matrices of size $2n \times (n+N)$. We will certify that a graph G on n vertices does not have any k-subgraph with excess ϵ if $\mathcal{A}$ accepts $C'(G)$. As in the proof of Theorem 6, for suitable choices of k, ϵ and δ this will yield a contradiction with Hypothesis 1. We now fill in the remaining details.

Instead of setting $\alpha = 3/4$ and $\beta = 1/3$ we now set $\alpha = \epsilon_0$; in order to satisfy the constraints of Lemma 4 we need to take β in the interval $]\epsilon_0, \frac{1}{2}(\frac{1}{2}+\epsilon_0)[$. For reasons that will soon become clear we pick a β which also satisfies the additional constraint $\beta < 2\epsilon_0$, and we set as usual $k = n^\alpha$ and $\epsilon = 1/n^\beta$. It remains to show that we can use $\mathcal{A}$ to certify that most graphs on n vertices do not have any k-subgraph with excess ϵ.

On the one hand, as explained in the proof of Theorem 6 for most graphs $C(G)$ satisfies the RIP of order k with parameter $\delta = c' n^{\alpha-\beta-1/2}$ (the constant c' is chosen as in that proof). Remember also that B_n satisfies the RIP of order k with parameter $n^{-\epsilon_0}$. Due to the choice $\beta < 2\epsilon_0$ we have $n^{-\epsilon_0} < \delta$ for n large enough, so by Lemma 6 for most graphs G the matrix $C'(G)$ will satisfy the RIP of order k with parameter δ.

On the other hand, we have seen in the proof of Theorem 6 that if n is large enough and G has a k-subgraph with excess ϵ then its RIP parameter is larger than $\lambda\delta$. By Lemma 6 the same is true of $C'(G)$. $\square$

We can also give a hardness result for rectangular matrices based on Hypothesis 2. In fact we need a randomized version of this hypothesis which we state explicitly below.

Hypothesis 3 (Randomized version of Hypothesis 2) *There is a constant* $\kappa \in]0, 1[$ *such that the following holds.*

Fix two arbitrary constants $\alpha \in]0, 1[$ *and* $\gamma > 0$ *with* $\alpha \leq \kappa$ *and* $\alpha\gamma \leq \kappa$. *Set* $\epsilon(n) = \gamma/\sqrt{n}$ *and* $k(n) = \alpha n$ *(again we only consider values of* n *such that* $k(n) \in \mathbb{N}$*). Then no polynomial time* randomized *algorithm can certify that most graphs on* n *vertices do not have any* $k(n)$*-subgraph with excess* $\epsilon(n)$*.*

In the above hypothesis, the hypothetical randomized algorithm would have to satisfy the following properties:

(i) For most graphs G on n vertices, with probability at least (say) 3/4 the algorithm certifies that G does not contain any $k(n)$-subgraph with excess $\epsilon(n)$.

(ii) For all graphs G on n vertices, if G is certified then it is always true (with probability 1) that G does not contain any $k(n)$-subgraph with excess $\epsilon(n)$.

Note that the probability bounds in (i) and (ii) refer to the *internal* coin tosses of the algorithm.

Theorem 9 *Fix any constant $\delta_0 < \kappa/2$, where κ is the constant from Hypothesis 3. Assuming this hypothesis, no polynomial time algorithm can distinguish between matrices which satisfy the RIP of order k with parameter δ_0 and matrices which do not satisfy the RIP of order k with parameter $\kappa/2$.*

Moreover, polynomial-time distinction between these two cases remains impossible even for matrices of size $2n \times 100n$.

Again, only the second part of the theorem is new.

Proof. As in the proof of Theorem 8 we construct from a graph G a matrix of the form

$$C'(G) = \begin{pmatrix} C(G) & 0 \\ 0 & B_n \end{pmatrix}.$$

For B_n, instead of of the deterministic construction from [6, 7] we will use a $n \times 99n$ random matrix given by Theorem 2. As before, we will certify that G does not have any k-subgraph with excess ϵ if the hypothetical distinguishing algorithm accepts $C'(G)$. This will yield a contradiction with the randomized version of Hypothesis 2.

In the proof of Theorem 7 we showed how to set the various parameters so that $C(G)$ does not satisfy the RIP of order k with parameter 1/2 in the case where G has a k-subgraph with excess ϵ. In this case, by Lemma 6 it is also true that $C'(G)$ does not satisfy the RIP of order k with parameter 1/2. Therefore we will never certify a graph which contains a k-subgraph with excess ϵ (note that this holds irrespective of the choice of B_n).

We now turn our attention to the other case. By Theorem 2, with probability approaching 1 as $n \to +\infty$ matrix B_n will satisfy the RIP of order k with parameter δ_0 if the condition

$$n \geq C\delta_0^{-2} k \log(eN/k). \tag{7}$$

is satisfied. Since $k = \alpha n$ and $N = 99n$ this is equivalent to $\delta_0^2 \geq C\alpha \log(99e/\alpha)$. Note that the right-hand side of this expression goes to 0 with α, but the left-hand side is constant. Condition 7 will therefore be satisfied for any small enough α. But recall from the proof of Theorem 7 that for small enough α we can set the other parameters so that $C(G)$ satisfies the RIP of order k with parameter δ_0 for most graphs G. With such a setting of parameters, we conclude from Lemma 6 that $C'(G)$ satisfies the RIP of order k with parameter δ_0 for most graphs and most choices of the random matrix B_n. □

The constant 100 in Theorem 9 can be replaced by any other constant. Note also that the hypothetical polynomial-time algorithm in this theorem remains deterministic: it is only the (hypothetical) algorithm for graph certification which is randomized. It is clear, however, that Theorem 9 remains true for randomized algorithms with one-sided error.

Note that depending on the value of the constant κ in Hypothesis 3, Theorem 9 does not rule necessarily rule out the existence of a polynomial time algorithm for deciding whether a RIP parameter is smaller than $\sqrt{2}-1$ (recall from the introduction that this is good enough for applications to compressed sensing). Such an algorithm is ruled out if $\sqrt{2}-1 \in]\kappa/2, \kappa[$, but could in principle exist if $\kappa < \sqrt{2}-1$. [1]

Likewise, our hardness results do not rule out the existence of a polynomial-time algorithm distinguishing between matrices with a very small RIP parameter and matrices with a RIP parameter larger than say 0.1. Here *very small* means as in Theorem 6 that the RIP parameter goes to 0 as $n \to +\infty$. If convergence to 0 is not too fast then we could still use such a weak distinguishing algorithm for certifying most random matrices. The following table gives a summary of our hardness results.

Hardness Results				
k	(k,δ_1) vs. (k,δ_2) - hard	**Result**	**Assumptions**	**Dimensions** ($n \times N$)
$n^{3/4}$	$\delta_1 = c_1 n^{-1/12}$, $\delta_2 = \lambda c_1 n^{-1/12}$ $\forall \lambda > 1$	Theorem 6	Hypothesis 1	$n \times n$
$n^{1/3}$	$\delta_1 = c_1 n^{-4/15}$, $\delta_2 = \lambda c_1 n^{-4/15}$ $\forall \lambda > 1$	Theorem 6	Hypothesis 1	$n \times n$
$\Theta(n)$	$\delta_1 = \kappa/2$, $\delta_2 = \kappa$	Theorem 7	Hypothesis 2	$n \times n$
$n^{3/4}$	$\delta_1 = c_2 n^{-1/12}$, $\delta_2 = \lambda c_2 n^{-1/12}$ $\forall \lambda > 1$	Theorem 8	Hypothesis 1	$n \times n^{1+\epsilon_0}$ for $\epsilon_0 > 0$
$\Theta(n)$	$\delta_1 = \kappa/2$, $\delta_2 = \kappa$	Theorem 9	Hypothesis 3	$n \times 50n$

Table 1: We say that a matrix Φ has the (k,δ)-RIP iff $(1-\delta) \leq \|\Phi x\|^2 \leq (1+\delta)$ for every k-sparse unit vector x. By (k,δ_1) vs. (k,δ_2)-hard we abbreviate the following: no polynomial time algorithm can distinguish matrices Φ that satisfy the (k,δ_1)-RIP from matrices that do not satisfy the (k,δ_2)-RIP. The constants c_1, c_2 depend on λ.

9 Acknowledgements

We thank Eden Chlamtáč for many useful discussions. In particular, Eden suggested to study the complexity of the RIP for square matrices, and helped refute an earlier version of Hypothesis 2.

[1] From Section 6 we know that $\kappa \leq \sqrt{5}/3 \simeq 0.745$, so in any case we have $\kappa/2 < \sqrt{2}-1$.

References

[1] M. Alekhnovich. More on Average Case vs Approximation Complexity. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 298–307, 2003.

[2] N. Alon, O. Goldreich, J. Hastad, and R. Peralta. Simple Construction of Almost k-wise Independent Random Variables. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 544–553, 1990.

[3] N. Alon, M. Krivelevich, and B. Sudakov. Finding a Large Hidden Clique in a Random Graph. *Random Struct. Algorithms*, 13:457–466, 1998.

[4] N. Alon, M. Krivelevich, and V. Vu. On the Concentration of Eigenvalues of Random Symmetric Matrices. *Israel Journal of Mathematics*, 131:259–267, 2002.

[5] A. Bhaskara, M. Charikar, E. Chlamtáč, U. Feige, and A. Vijayaraghavan. Detecting High log-densities: an $\mathcal{O}(n^{1/4})$-approximation for densest k-subgraph. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 201–210, 2010.

[6] J. Bourgain, S. J. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova. Explicit Constructions of RIP Matrices and Related Problems. Available at arxiv:1008.4535, August 2010.

[7] J. Bourgain, S. J. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova. Breaking the k^2 Barrier for Explicit RIP Matrices. In *Proceedings of the Symposium on Theory of Computing (STOC)*, 2011. Conference version of [6].

[8] E. J. Candès. The Restricted Isometry Property and its Implications for Compressed Sensing. *Comptes Rendus Mathematique*, 346(9-10):589 – 592, 2008.

[9] E. J. Candès, J. K. Romberg, and T. Tao. Stable Signal Recovery from Incomplete and Inaccurate Measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[10] E. J. Candès and T. Tao. Decoding by Linear Programming. *Information Theory, IEEE Transactions on*, 51(12):4203 – 4215, 2005.

[11] A. d'Aspremont, F. Bach, and L. El Ghaoui. Optimal Solutions for Sparse Principal Component Analysis. *J. Mach. Learn. Res.*, 9:1269–1294, 2008.

[12] A. d'Aspremont and L. El Ghaoui. Testing the Nullspace Property using Semidefinite Programming. Available at arxiv:0807.3520, July 2008.

[13] R. A. DeVore. Deterministic Constructions of Compressed Sensing Matrices. *J. Complex.*, 23:918–925, 2007.

[14] U. Feige. Relations between Average Case Complexity and Approximation Complexity. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 534–543, 2002.

[15] Z. Füredi and J. Komlós. The Eigenvalues of Random Symmetric Matrices. *Combinatorica*, 1:233–241, 1981.

[16] A. Juditsky and A. Nemirovski. On Verifiable Sufficient Conditions for Sparse Signal Recovery via ℓ_1 Minimization. Available at arxiv:0809.2650, September 2008.

[17] B. S. Kashin. The Diameters of Octahedra. *Uspekhi Mat. Nauk*, 30:251–252, 1975.

[18] S. Khot. Ruling out PTAS for Graph min-bisection, Dense k-subgraph, and Bipartite Clique. *SIAM J. Comput.*, 36:1025–1071, 2006.

[19] M. Ledoux. On Talagrand's Deviation Inequalities For Product Measures. *ESAIM Probab. Stat.*, 1:63–87, 1996.

[20] M. Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.

[21] R. Vershynin. Introduction to the Non-asymptotic Analysis of Random Matrices. In: Compressed Sensing: Theory and Applications, eds. Y. Eldar and G. Kutyniok. Cambridge University Press, to appear. Available at http://www-personal.umich.edu/ romanv/papers/non-asymptotic-rmt-plain.pdf, 2010.

[22] V. Vu. Spectral Norm of Random Matrices. *Combinatorica*, 27:721–736, 2007.

Appendix

Proof of Theorem 4 We recall a bound on the expectation of a symmetric Bernoulli matrix [22, Eqn. 1].

Lemma 7 *Let A be a symmetric matrix whose off-diagonal entries are i.i.d. from $\{\pm 1\}$ with equal probability and having zero diagonal entries. There exist positive constants δ, c_1 and c_2 so that for $k = c_1 n^\delta$*

$$\mathbb{E}\boldsymbol{tr}\left(A^k\right) \leq c_2 n (2\sqrt{n})^k. \tag{8}$$

By definition of the operator norm, we know that $\|A\| := \sup_{\|x\|=1} \|Ax\|$ for any matrix A. Decompose any unit vector as $x = \alpha \mathbf{1}/\sqrt{n} + \beta y$, where $\mathbf{1}$ is the all-ones vector, y is the unit vector with $y \perp \mathbf{1}$ and $\alpha^2 + \beta^2 = 1$.

$$\begin{aligned}
\mathbb{E}\left\|\widehat{A}\right\| &= \mathbb{E}\sup_{\|x\|=1}\left\|\widehat{A}x\right\| = \mathbb{E}\sup_{\|y\|=1,\ y\perp\mathbf{1}}\ \sup_{\alpha^2+\beta^2=1}\left\|\widehat{A}(\alpha\frac{\mathbf{1}}{\sqrt{n}}+\beta y)\right\| \\
&\leq \mathbb{E}\sup_{\|y\|=1,\ y\perp\mathbf{1}}\ \sup_{\alpha^2+\beta^2=1}|\alpha|\left\|\widehat{A}\frac{\mathbf{1}}{\sqrt{n}}\right\| + |\beta|\left\|\widehat{A}y\right\| \leq \mathbb{E}\sup_{\|y\|=1,\ y\perp\mathbf{1}}\sqrt{\left\|\widehat{A}\frac{\mathbf{1}}{\sqrt{n}}\right\|^2 + \left\|\widehat{A}y\right\|^2} \\
&= \mathbb{E}\sqrt{\left\|\widehat{A}\frac{\mathbf{1}}{\sqrt{n}}\right\|^2 + \sup_{\|y\|=1,\ y\perp\mathbf{1}}\left\|\widehat{A}y\right\|^2} \leq \sqrt{\mathbb{E}\left\|\widehat{A}\frac{\mathbf{1}}{\sqrt{n}}\right\|^2 + \sup_{\|y\|=1,\ y\perp\mathbf{1}}\left\|\widehat{A}y\right\|^2} \\
&\leq \sqrt{\mathbb{E}\left\|\widehat{A}\frac{\mathbf{1}}{\sqrt{n}}\right\|^2 + \mathbb{E}\left\|A/2\right\|^2}
\end{aligned} \tag{9}$$

where we used the triangle's inequality, Cauchy-Schwarz's inequality, Jensen's inequality and the fact that $\mathbf{J}y = 0$ for every $y \perp \mathbf{1}$. Now we bound each of the two terms on the right-hand side of (9) separately; the first term is straight-forward to bound. Indeed,

$$\begin{aligned}
\left\|\widehat{A}\frac{\mathbf{1}}{\sqrt{n}}\right\|^2 &= \left\|A\frac{\mathbf{1}}{2\sqrt{n}} + a\mathbf{1}\right\|^2 = \left\|A\frac{\mathbf{1}}{2\sqrt{n}}\right\|^2 + 2\left\langle A\frac{\mathbf{1}}{2\sqrt{n}},\ a\mathbf{1}\right\rangle + a^2 n \\
&= \left\|A\frac{\mathbf{1}}{2\sqrt{n}}\right\|^2 + (a/\sqrt{n})\sum_{i,j} A_{i,j} + a^2 n = \frac{1}{4n}\sum_i \left(\sum_j A_{ij}\right)^2 + (a/\sqrt{n})\sum_{i,j} A_{i,j} + a^2 n \\
&= \frac{1}{4n}\sum_i \left(\sum_j A_{ij}^2 + 2\sum_{l<k} A_{il}A_{ik}\right) + (a/\sqrt{n})\sum_{i,j} A_{i,j} + a^2 n
\end{aligned}$$

Now, take expectation and recall that $\mathbb{E}A_{ij} = 0$ and $\mathbb{E}A_{ij}^2 = 1$. Since the entries of A are independent, we obtain the inequality:

$$\mathbb{E}\left\|\widehat{A}\frac{\mathbf{1}}{\sqrt{n}}\right\|^2 \leq (a^2 + 1/4)n. \tag{10}$$

To bound the second term, we use Wigner's trace method (Lemma 7). For k as in Lemma 7,

$$\mathbb{E}\|A\|^2 = \mathbb{E}\sqrt[k]{\|A^{2k}\|} \leq \mathbb{E}\sqrt[k]{\mathbf{tr}\left(A^{2k}\right)} \leq \sqrt[k]{\mathbb{E}\mathbf{tr}\left(A^{2k}\right)} \leq 4n2^{\mathcal{O}(\log(n)/n^\delta)}.$$

where we used that $\left\|A^{2k}\right\| \leq \mathbf{tr}\left(A^{2k}\right)$, Jensen's inequality and Lemma 7. The above discussion implies that

$$\mathbb{E}\|A/2\|^2 \leq n(1 + o(1)). \tag{11}$$

Combining (9),(10) and (11), it follows that

$$\mathbb{E}\left\|\widehat{A}\right\| \leq \sqrt{n}\sqrt{a^2 + 5/4 + o(1)}.$$

This completes the proof of the first part of Theorem 4. Now, we describe a tail bound for a convex function of random signs. This result was established by Ledoux [19]; see also [20, § 5.2] for a discussion of concentration in product spaces.

Theorem 10 *Suppose $f : \mathbb{R}^m \to \mathbb{R}$ is a convex function that satisfies the Lipschitz bound*

$$|f(x) - f(y)| \leq L\,\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^m.$$

Let $r \in \{\pm 1\}^m$ be a random sign vector. For all $\epsilon \geq 0$,

$$\Pr\left[f(r) \geq \mathbb{E}f(r) + L\epsilon\right] \leq \exp(-\epsilon^2/8).$$

Let $f : \mathbb{R}^{n(n-1)/2} \to \mathbb{R}$. Consider the bijection between the set of symmetric matrices of size n having zero diagonal and the set of vectors in $\mathbb{R}^{n(n-1)/2}$ via the natural vectorization function $\mathrm{vec} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n(n-1)/2}$, i.e., vectorize the strict lower triangular part of its argument matrix. For every vector $v \in \mathbb{R}^{n(n-1)/2}$, define $f(v) := \left\|\frac{1}{2}\mathrm{vec}^{-1}(v) + \frac{a}{\sqrt{n}}\mathbf{J}\right\|$. By definition, f is convex. Let's compute its Lipschitz constant; let A, B be symmetric matrices of size n with zero diagonals, then $|f(\mathrm{vec}(A)) - f(\mathrm{vec}(B))| = \left|\left\|\frac{1}{2}A + \frac{a}{\sqrt{n}}\mathbf{J}\right\| - \left\|\frac{1}{2}B + \frac{a}{\sqrt{n}}\mathbf{J}\right\|\right| \leq \left\|\frac{1}{2}(A - B)\right\| \leq 1/2\,\|A - B\|_{\mathrm{F}} \leq \|\mathrm{vec}(A) - \mathrm{vec}(B)\|$ where $\|A\|_{\mathrm{F}}^2 = \sum_{i,j} A_{ij}^2$ (the inequality $\|A\| \leq \|A\|_{\mathrm{F}}$ always holds). Hence the Lipschitz constant of f is one, which concludes the proof of the "moreover" part of Theorem 4.